Un error de Microsoft expone correos electrónicos confidenciales a la herramienta de inteligencia artificial Copilot

Hace 2 días

Liv McMahon
reportero de tecnología

Imágenes Getty

Microsoft ha reconocido un error que provocó que su asistente de trabajo de inteligencia artificial accediera y resumiera por error los correos electrónicos confidenciales de algunos usuarios.

El gigante tecnológico ha impulsado Microsoft 365 Copilot Chat como una forma segura para que los lugares de trabajo y su personal utilicen su chatbot de inteligencia artificial generativa.

Pero dijo que un problema reciente provocó que la herramienta mostrara información a algunos usuarios empresariales desde los mensajes almacenados en sus borradores y carpetas de correo electrónico enviado, incluyendo aquellos marcados como confidenciales.

Microsoft dice que ha lanzado una actualización para solucionar el problema y que "no proporcionó a nadie acceso a información que no estaba autorizado a ver".

Sin embargo, algunos expertos advirtieron que la velocidad a la que las empresas compiten para agregar nuevas funciones de IA significaba que este tipo de errores eran inevitables.

Copilot Chat se puede utilizar dentro de programas de Microsoft como Outlook y Teams, para correos electrónicos y funciones de chat, para obtener respuestas a preguntas o resumir mensajes.

"Identificamos y solucionamos un problema por el cual Microsoft 365 Copilot Chat podía devolver contenido de correos electrónicos etiquetados como confidenciales creados por un usuario y almacenados en sus Borradores y Elementos enviados en el escritorio de Outlook", dijo un portavoz de Microsoft a BBC News.

"Si bien nuestros controles de acceso y políticas de protección de datos se mantuvieron intactos, este comportamiento no cumplió con la experiencia Copilot prevista, que está diseñada para excluir el contenido protegido del acceso a Copilot", agregaron.

"Se ha implementado una actualización de configuración en todo el mundo para clientes empresariales".

El error fue reportado por primera vez por el medio de noticias tecnológicas **Bleeping Computer** , que dijo que había visto una alerta de servicio que confirmaba el problema.

Citó un aviso de Microsoft que decía que "los mensajes de correo electrónico de los usuarios con una etiqueta confidencial aplicada están siendo procesados incorrectamente por el chat de Microsoft 365 Copilot".

El aviso agregó que una pestaña de trabajo dentro de Copilot Chat había resumido los mensajes de correo electrónico almacenados en los borradores y carpetas de enviados de un usuario, incluso cuando tenían una etiqueta de confidencialidad y una política de prevención de pérdida de datos configurada para evitar el intercambio de datos no autorizado.

Los informes sugieren que Microsoft se dio cuenta del error por primera vez en enero.

Su aviso sobre el error también se compartió en un panel de soporte para trabajadores del NHS en Inglaterra, donde la causa raíz se atribuye a un "problema de código".

Una sección del **aviso en el sitio de soporte de TI del NHS** implica que se ha visto afectado.

Pero le dijo a BBC News que el contenido de cualquier borrador o correo electrónico enviado y procesado por Copilot Chat permanecería con sus creadores y que la información de los pacientes no ha sido expuesta.

“Se producirán fugas de datos”

Las herramientas de inteligencia artificial empresarial, como Microsoft 365 Copilot Chat (disponible para organizaciones con una suscripción a Microsoft 365), a menudo tienen controles y protecciones de seguridad más estrictos para evitar compartir información corporativa confidencial.

Pero para algunos expertos, la cuestión aún resalta los riesgos de adoptar herramientas de IA generativa en ciertos entornos de trabajo.

Nader Henein, analista de protección de datos y gobernanza de IA en Gartner, dijo que "este tipo de errores son inevitables", dada la frecuencia con la que se lanzan "capacidades de IA nuevas y novedosas".

Dijo a BBC News que las organizaciones que utilizan estos productos de IA a menudo carecen de las herramientas necesarias para protegerse y gestionar cada nueva función.

"En circunstancias normales, las organizaciones simplemente desactivarían la función y esperarían hasta que la gobernanza se pusiera al día", dijo Henein.

"Desafortunadamente, la cantidad de presión causada por el torrente de propaganda infundada sobre IA hace que eso sea casi imposible", agregó.

El experto en seguridad cibernética, profesor Alan Woodward de la Universidad de Surrey, dijo que esto demuestra la importancia de hacer que dichas herramientas sean privadas por defecto y sólo con suscripción voluntaria.

"Es inevitable que estas herramientas tengan errores, sobre todo porque avanzan a una velocidad vertiginosa, así que, aunque la fuga de datos no sea intencionada, ocurrirá", declaró a BBC News.